Remarks

Reconsideration of the above referenced application in view of the enclosed remarks is requested. Applicants note with appreciation that the Examiner has allowed Claims 3-5, 27-31, 35-37 and 40-46. Claims 1-2, 6-26, 32-34, and 38-39 have been rejected. Existing Claims 1-46 remain in the application.

ARGUMENT

Claims 1, 2, 6, 7, 9, 10, 13-16, 18, 19, 21, 23, 24 and 32 are rejected under 35 U.S.C. § 103(a) as being unpatentable over ("Hybrid Language Models and Spontaneous Legal Discourse", 4th International Conference on Spoken Language, October 1996) (hereafter, Kenne et al.) in view of Hsu et al. (U.S. Patent 5,677,991 A) (hereafter, Hsu et al.) and further in view of Guerreri (U.S. Patent 5,189,727 A) (hereafter, Guerreri). This rejection is respectfully traversed and Claims 1, 2, 6, 7, 9, 10, 13-16, 18, 19, 21, 23, 24 and 32 are believed allowable based on the following discussion.

Regarding Claims 1, 13, 18 and 23, the Examiner asserts that Kenne et al. teaches selecting a recognizer; receiving an input stream; deriving selection information, wherein the selection information includes performance-related information; using the selection information to select results from at least one enabled recognizer. The Examiner admits that Kenne et al. does not mention applications, but asserts that this element is taught by Hsu et al. Claims 1, 13, and 18 were previously amended to require that a recognizer is enabled based upon an expected future performance of the recognizer. The Examiner has previously indicated that this limitation is directed toward allowable subject matter. Therefore Claims 1, 13, 18 and their progeny are allowable.

With respect to the newly cited reference, the Examiner asserts that Guerreri teaches that the "future selection" technique is the same as Applicants' recited future performance. This comparison is erroneous. Guerreri teaches a method and apparatus for

> "language and speaker recognition where an initial learning phase creates histograms for each of the languages to be recognized. A first pass enters a number of

> samples of speech, and at each predetermined instant of time, each sample of speech is Fast Fourier Transformed (FFT) to create a spectrum showing frequency content of the speech at that instant of time (a spectral vector). The frequency content is compared with frequency contents which have been previously stored. If the current spectral vector is close enough to a previously stored spectral vector, a weighted average between the two is formed, and a weight indicating frequency of occurrence is incremented. If the current value is not similar to one which has been previously stored, it is stored with an initial weight of "1".

While Guerreri stores previously collected data, at no time is it taught or suggested that a recognizer is enabled based upon an expected future performance of the recognizer. Guerreri teaches of method of using past data within the algorithm of one recognizer, not a future performance expectation which allows the selection among a plurality of recognizers. The future selection discussed by Guerreri is a selection of a language from a set of languages being matched by the pattern recognition algorithm. Guerreri uses an algorithm to select a language using pattern recognition techniques in a single recognizer. In contrast, Applicants recite a system and method for selecting one of a plurality of recognizers and enabling a recognizer based on an expected future performance. Guerreri does not teach or suggest a system with multiple recognizers where a recognizer is enabled based upon expected future performance. Thus, not only is there no motivation to combine the teachings of Guerreri with the teachings of Kenne et al. and Hsu et al., doing so will not result in Applicants' claimed invention. Nor will combining the references result in a system that enables a recognizer based on expected future performance. Selecting a language is not similar to selecting an enabled recognizer based on expected future performance. Applicants respectfully request that the Examiner allow these claims to issue at the earliest possible time.

With regard to Claim 23, Applicants require using the enabling information to select an enabled recognizer. This limitation is neither taught nor suggested by Kenne et al. or Hsu et al. The cited reference in Guerreri is not relevant to the limitations recited in Claim 23, as discussed above and therefore this rejection is improper. Kenne et al. does not teach to enable recognizers. The models used in Kenne et al. are turned on one at a time. The local perplexity determines the selection of the model (or recognizer). Kenne et al. teaches the training of three models, Both, L+W, and Hybrid. In operation, the Kenne et al. speech recognizer takes an input stream and applies it to one model, based on the source (lawyer or

witness). In contrast, Applicants' claimed invention selects which recognizers to send the input stream to, either in parallel or in sequence, using enabling information. More than one recognizer may be enabled. The results are selected from one of the enabled recognizers based on selection information in the predictor. By selectively enabling recognizers, one or more recognizers may be omitted from the selection of the results. This is important to override a recognizer when it may otherwise be the one selected to have the best results. Enabling information, as defined by Applicants is at least described on page 4 of the specification.

The effect of disabling certain recognizers, i.e., those not enabled, is not the same as selecting one specific model (recognizer) as is taught by Kenne et al. Kenne et al. has no mechanism for taking results from more than one recognizer at the same time and predicting the best results. Therefore, Claims 23 is believed allowable, and should be allowed to issue.

Regarding Claim 2, the Examiner asserts that Kenne et al. teaches the feature that causes a recognizer to be selected that is different from the recognizer used in a previous interaction. The Examiner admits that Kenne et al. and Guerreri do not teach that the selection information is updated, but asserts that this is taught by Hsu et al. by setting a reference score as a baseline that is necessary to provide further evaluation. This rejection is respectfully traversed based on the foregoing and following discussion. Claim 2 is allowable because it is dependent on a claim with allowable subject matter. Specifically, at least, the cited references do not teach or suggest using the selection information to select results from at least one enabled recognizer, wherein a recognizer is enabled based upon an expected future performance of the recognizer, as recited in Claim 1.

Regarding Claim 6 and 24, the Examiner asserts that Results section of Kenne et al. discloses the feature that the enabling information, and consequently the performance-related information, comprises at least one type of information from the group comprised of: channel characteristics, user information, contextual information, dialog state, recognizer costs and performance history. This rejection is respectfully traversed based on the foregoing and following discussion. Neither Kenne et al., nor Hsu et al. teach deriving, analyzing or using enabling information, especially recognizer costs. Applicants describe that one element of performance-related information is a quantitative analysis of the costs of using a particular

recognizer, in either financial or computational terms. This information is part the enabling information. Applicants use the enabling information to determine which recognizers to enable. Only the enabled recognizers process the input stream. The results of the enabled recognizers are then selected based on performance-related information. This is neither taught nor suggested by the cited references. Further, Claims 6 and 24 are dependent on allowable Claims 1 and 23, respectively. Thus, Claims, 6 and 24 are believed allowable.

Regarding Claims 7 and 14, the Examiner asserts that Kenne et al. teaches deriving the selection information further comprises analyzing the input stream for channel characteristics. Claims 7 and 14 are dependent on a Claim with allowable subject matter. Therefore, Claims 7 and 14 are allowable. Further, with respect to Claims 7 and 14, Kenne et al. does not teach analyzing the input stream for channel characteristics as defined by Applicants. Kenne et al. merely determines on which track the transcript to be recognized is located and selects a model accordingly. Kenne et al does not "determine the audio characteristics of the channel, for example, determining if the cellular or landline communication networks are in use." Kenne et al. does not derive or analyze background noise and signal strength or other channel characteristics. The analysis as described in the specification may determine characteristics of the communication device, for example determining if a speaker-phone or a wireless handset are in use. Network-based information services such as CallerID in conjunction with a local or network-based database mapping calling number to channel and device characteristics may be utilized for similar effect. The cited references do not teach or suggest analyzing the input stream for channel characteristics to derive selection information.

Regarding Claims 9, 15 and 19, the Examiner asserts that Kenne et al. discloses receiving contextual information associated with the input stream by virtue of distinguishing between Lawyers and Witnesses. Claims 9, 15 and 19 are dependent on a Claim with allowable subject matter. Therefore, Claims 9, 15, and 19 are allowable. Kenne et al. merely discloses being able to distinguish between two tracks of information to determine which recognition model to use. Applicants' claimed invention requires that selection information comprises receiving contextual information associated with the input stream. Contextual information is defined in the specification as originally filed as that information related to the

environment around the input stream, including characteristics of the user and information derived from the call using network services such as CallerID. This information may be obtained dynamically or may be predetermined. Contextual information may include gender, age, ethnicity, whether the speaker speaks the language of the recognizers as a first language, among other personal information about the user. Also, the channel and device characteristics may be included in the contextual information. Kenne et al. does not derive contextual information as defined by the Applicants.

Regarding Claims 10 and 16, the Examiner asserts that Kenne et al. teaches the feature of receiving recognizer information from the enabled recognizers to be used in the selection information with respect to the switching described in Section 4 (Results) of Kenne et al. Kenne et al. does not receive recognizer information from the enabled recognizers to be used in the selection information. First, Kenne et al. does not teach enabled recognizers, as defined by Applicants. Kenne et al. does not teach or disclose recognition information as defined by Applicants, and as discussed above. Thus, Kenne et al, does not teach or disclose receiving recognizer information from the enabled recognizers to be used in selection information. Further, Claims 10 and 16 are allowable as being dependent upon a claim with allowable subject matter.

Regarding claim 21, the Examiner asserts that Kenne et al. teaches that the predictor is operable to select a recognizer based upon the converted stream. Claim 21 is allowable as being dependent upon a claim with allowable subject matter.

Regarding Claim 32, the Examiner asserts that Claim 32 is set forth with the same limits as Claim 18. Claim 32 is directed toward a method dependent on Claim 9. Claim 18 is directed toward a system. Thus, the Examiner's rejection is improper. Further, the Examiner asserts that Guerreri teaches a feature *wherein contextual information comprises information from at least one item of information derived from the set of information comprising information related to the environment around the input stream, characteristics of a user generating the input stream, information derived from a call using network services, gender, age, ethnicity, information relating to the user's first (native) language, personal information about the user, channel characteristics and device characteristics*. Guerreri notes in the Background section that:

> "There are many applications where it is desirable to determine an aspect of spoken sounds. This aspect may include identifying a language being spoken, identifying a particular speaker, identifying a device, such as a helicopter or airplane and a type of the device, and identifying a radar signature, for instance. For instance, a user may have a tape recording of information, which the user needs to understand. If this information is in a foreign language, it may be required to be translated. However, without knowing what language the information is in, it will be difficult for the user to choose a proper translator."

The mere act of mentioning that there are different aspects to spoken sounds does not teach the limitations of the claimed invention. Guerreri does not teach or suggest *deriving the selection information further comprises receiving contextual information associated with the input stream.* Nor does Guerreri teach that the items listed in Col. 1, lines 19-30 are contextual information used to derive selection information, or anything else. Thus, Claim 32 is allowable as being dependent on an allowable claims and because the limitations are not taught or suggested by the cited references.

Claim 8 is rejected under 35 U.S.C. § 103(a) as being unpatentable over Kenne et al. in view of Hsu et al. and further in view of Guerreri and further in view of Waibel et al. (U.S. Patent 5,712,957 A) (hereafter Weibel '957). The Examiner admits that neither Kenne et al. nor Hsu et al. teach separate input devices. The Examiner asserts that Waibel '957 discloses different inputs and uses information which corresponds to analyzing the input stream for device characteristics. This rejection is respectfully traversed and Claim 8 is believed allowable based on the foregoing and following discussion.

The Examiner references elements 23 and 24 in Figure 1 of Weibel '957 to show multiple input streams. This reference is erroneous because element 24 is not an input stream that could be recognized by a speech recognizer. It is not speech or in the same category of input as the other input stream. In Col. 4, lines 53-60, Weibel '957 describes 24 as a "touch sensitive pad" or other input transducer. Weibel '957 describes translating this input with a handwriting recognition engine or other device entry. This input is to be used to assist with correction and repair of the module (recognizer). It is not an alternative input stream to be analyzed for characteristics that will derive selection information derived from the device characteristics. It is an additional input stream that is used to correct the first input stream

with related information. Further, Claim 8 is allowable based on its dependence from a claim with allowable subject matter.

Claims 11-12, 17 and 25-26 are rejected under 35 U.S.C. § 103(a) as being unpatentable over Kenne et al. in view of Hsu et al. and further in view of Guerreri and further in view of Kundu (U.S. Patent 5,924,066 A) (hereafter, Kundu). This rejection is respectfully traversed based on the foregoing and following discussion.

Regarding Claims 11, 17 and 25, the Examiner admits that Kenne et al. and Hsu et al. do not disclose feedback. The Examiner asserts that Kundu teaches this with classifying a speech signal that receiving feedback and including the feedback in the selection information is disclosed. Kundu does not teach or disclose feedback as described by Applicants. Kundu teaches using a neural net to train the recognizer with data. Kundu describes a 3-layer neural net as follows:

> "The learning law for the perceptron 40 is a simple error feedback. The network learns the associations between input and output patterns by being exposed to many lessons. The weights are adjusted until the desired target output is produced. This weight adaptation is referred to as error backpropagation learning law."

Kundu discloses using an error backpropagation algorithm in the neural net. This does not generate "feedback." A neural net, as described by Kundu, cannot update recognizer information as defined in Applicants' claimed invention, but it merely provides a weighting for the input factors and provides a likely output. The output of the middle layer of the neural net, which is the heart of the error backpropagation algorithm, is not human readable/comprehensible. It does not provide feedback which can be used in the selection information. The top layer (output) of the neural net provides the best guess at the result (i.e. speech components) and does not provide information about performance or other recognizer information. Applicants' feedback provides input to a predictor, or selection information, which is used to select the recognizer from a plurality of recognizers operating on the same input stream, either in parallel, or in sequence. Kundu teaches a neural net which uses an error backpropagation method to train one recognizer. Thus, the error backpropagation does not provide feedback which will result in Applicants' claimed invention.

Regarding Claims 12 and 26, the Examiner asserts that Kundu teaches that feedback is received from one of the group comprised of: off-line analysis, user feedback, and

feedback from the recognizer. The neural net as described by Kundu provides a "simple error feedback." Error feedback in the context of a neural net is error associated with the results as they correspond to the training data. When in recognition mode, rather than training mode, the error feedback is already programmed into the neural net and feedback with respect to off-line analysis, user feedback, and feedback from the recognizer is not provided. Thus, claims 12 and 26 are believed allowable.

Claims 20, 22, 38 and 39 are rejected under 35 U.S.C. § 103(a) as being unpatentable over Kenne et al. in view of Hsu et al. and further in view of Guerreri and further in view of Waibel et al. (U.S. Patent 5,855,000) (hereafter, "Weibel '000"). This rejection is respectfully traversed and Claims 20, 22, 38 and 29 are believed allowable based on the foregoing discussion as being dependent upon a claim with allowable subject matter. Further, with respect to Claims 38 and 39, Waibel '000 does not teach or suggest recognizer-based confidence values for enabled recognizers. Waibel '000 teaches a confidence score for a single recognition. In contrast, Applicants' claims recite a predictor that uses a recognizer-based confidence value. Page 6 of the specification as originally filed describes: *"the predictor mechanism determines, for each recognizer in the system and for each situation, a recognizer-based confidence value. This recognizer-based confidence value is the predictor's estimation of the accuracy of each recognizer in a particular situation."* The confidence value as taught by Waibel '000 would not result in Applicants' invention if combined with the other cited references. Further, Waibel '000 discusses the confidence values in the context requiring the input or speech to be repeated. At no time is the confidence value taught by Waibel '000 used to select a recognizer based on expected future performance of the recognizer.

Claims 33 and 34 are rejected under 35 U.S.C. § 103(a) as being unpatentable over Kenne et al. in view of Hsu et al. and further in view of Guerreri and further in view of Goldberg et al. (U.S. Patent 5,970,446 A) (hereafter, Goldberg et al.). This rejection is respectfully traversed based on the foregoing and following discussion. Goldberg et al. teaches using location information to select a noise model for a single recognizer. There is no motivation to combine Goldberg et al. with the other cited references to use these techniques to derive the selection information using contextual information associated with

the input stream. Goldberg et al. along with the other cited references do not teach or suggest using the selection information which has been derived using contextual information to select results from at least one enabled recognizer and where a recognizer is enabled based upon an expected future performance of the recognizer. Thus, Claims 33 and 34 are allowable.

CONCLUSION

In view of the foregoing, 1-2, 6-26, 32-34, and 38-39 should be allowed along with allowed Claims 3-5, 27-31, 35-37 and 40-46. Thus, Claims 1-46 are all in condition for allowance and should be allowed to issue at the earliest possible time. Please charge any shortage of fees in connection with the filing of this paper, including extension of time fees, to Deposit Account 50-0221 and please credit any excess fees to such account. If the Examiner has any questions, the Examiner is invited to contact the undersigned at (703) 633-6845. Early issuance of Notice of Allowance is respectfully requested.

Respectfully submitted,

Dated: 9/7/2004

Joni D. Stutman-Horn
Patent Attorney
Intel Corporation
Registration No. 42,173
(703) 633-6845

Intel Americas, Inc.
4030 Lafayette Center Drive
MS LF3
Chantilly, VA 20151